Parametric Multichannel Audio Coding: Synthesis of Coherence Cues

Christof Faller

***Abstract*—Parametric multichannel audio coding represents an audio signal as one single audio channel plus side information. The side information contains estimates of perceptually relevant differences between the original audio channels. Usually, time difference, level difference, and coherence cues are considered. These cues determine, to a large degree, the auditory spatial image that is perceived when playing back multichannel audio signals. Level difference and time difference synthesis is simple: Different gain factors and delays are applied to the sum signal in subbands for generating the different decoder output channels. However, it is not as obvious how coherence cues can be synthesized. Several heuristic methods for coherence synthesis were proposed previously. In this paper, we are proposing a systematic approach for coherence synthesis. The coherence that is measured in the encoder between a pair of channels is reproduced in the decoder. For that purpose, de-correlation filters modeling late reverberation with impulse responses of a length of several hundred milliseconds are used, resulting in the ability of the scheme to generate naturally sounding diffuse sound. A method for reducing the computational complexity of the scheme is presented. The results of a subjective test indicate that the proposed scheme achieves good audio quality. Furthermore, the scheme was compared to a previous scheme without multichannel coherence synthesis and performs significantly better for all items tested.**

***Index Terms*—Auditory spatial image, diffuse sound, late reverberation, parametric multichannel audio coding, spatial perception, surround.**

I. Introduction

RECENTLY, there has been a renewed interest in parametric stereo[1] and multichannel audio coding techniques. Some of the limitations of the original parametric stereo audio coding technique, *intensity stereo* (IS) [1], [2], have been overcome by using different filterbanks for coding of the audio waveform and for parametric stereo [3]. Most audio coders use a *modified discrete cosine transform* (MDCT) [4] for coding of audio waveforms. The advantages of using a different filterbank for parametric stereo are reduced aliasing [3] and more flexibility, such as the ability to efficiently synthesize not only intensities but also time delays and coherence cues [5] between the audio channels.

Fig. 1 shows a generic scheme of *binaural cue coding* (BCC) [5]–[7]. BCC for *natural rendering* [5] is a parametric multichannel audio coding technique. As indicated in Fig. 1, the input audio channels $x_c(n)$ $(1 \leq c \leq C)$ are downmixed to one single audio channel $s(n)$, denoted *sum signal*. Downmixing by means of addition of the input channels and more sophisticated techniques have been proposed [5]. *Inter-channel time difference* (ICTD), *inter-channel level difference* (ICLD), and *inter-channel coherence* (ICC) cues are estimated between pairs of channels as a function of time and frequency. The sum signal and side information (the estimated cues) are transmitted to the BCC decoder, which generates its output channels $\hat{x}_c(n)$ $(1 \leq c \leq C)$ such that ICTD, ICLD, and ICC between the channels approximate those of the original audio signal.

The described scheme is able to represent multichannel audio signals at a bitrate only slightly higher than what is required to represent a mono audio signal. This is so, because the estimated ICTD, ICLD, and ICC between a channel pair contain one to two orders of magnitude less information than an audio waveform.

This paper is organized as follows. In Section II, details about BCC are described, such as how ICTD, ICLD, and ICC are defined and estimated. Also, previous techniques for ICC synthesis are reviewed and the ICC synthesis proposed in this paper is motivated. In Section III, spatial audio playback and spatial hearing are discussed. Based on this discussion, Section IV motivates the use of ICTD, ICLD, and ICC by BCC for representing attributes of the auditory spatial image. The proposed stereo and multichannel BCC synthesis schemes are described in Section V. Section VI describes how to implement the proposed schemes in the frequency-domain for reduced computational complexity. The results of subjective audio quality evaluations of the proposed schemes are presented in Section VII. Conclusions are drawn in Section VIII.

II. Multichannel BCC

Frequency dependence of ICTD, ICLD, and ICC is considered in BCC by estimating these cues in a subband domain as is illustrated in Fig. 2. The time and frequency resolution with which BCC estimates the cues is perceptually motivated and discussed in Section IV. The following definitions are used for ICTD, ICLD, and ICC for corresponding subband signals $\tilde{x}_1(k)$ and $\tilde{x}_2(k)$ of two audio channels with time index k

- ICLD [dB]

$$\Delta L_{12}(k) = 10 \log_{10} \left(\frac{p_{\tilde{x}_2}(k)}{p_{\tilde{x}_1}(k)} \right) \tag{1}$$

where $p_{\tilde{x}_1}(k)$ and $p_{\tilde{x}_2}(k)$ are short-time estimates of the power of the signals $\tilde{x}_1(k)$ and $\tilde{x}_2(k)$, respectively.

Manuscript received December 18, 2003; revised October 30, 2004. The work for this paper was carried out at Agere Systems, Allentown, PA. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ravi P. Ramachandran.

The author is with the Audiovisual Communications Laboratory, EPFL Lausanne, 1015 Lausanne, Switzerland, (e-mail: christof.faller@epfl.ch).

Digital Object Identifier 10.1109/TSA.2005.854105

[1]The term "stereo" always refers to two-channel stereophony only.

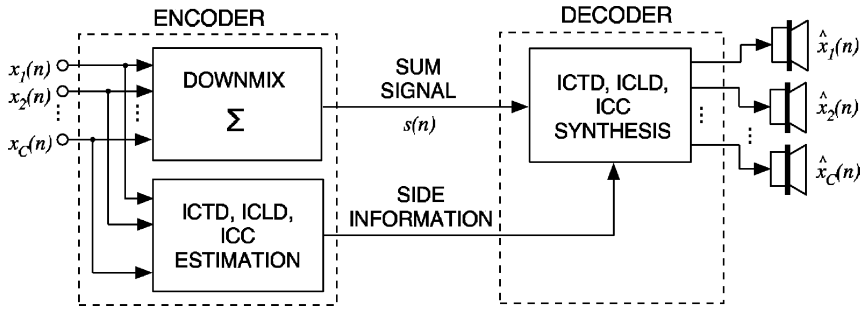

Fig. 1. Generic BCC scheme. A number of input signals are downmixed to one channel and transmitted to the decoder together with side information.

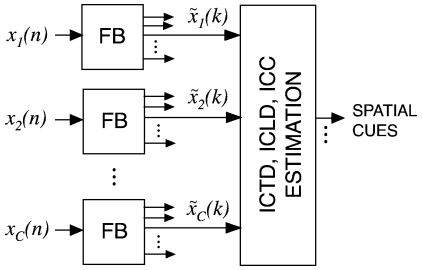

Fig. 2. Spatial cues, ICTD, ICLD, and ICC are estimated in a subband domain. The spatial cue estimation is applied independently to each subband.

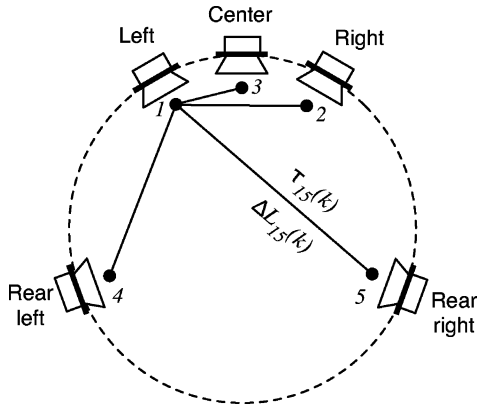

Fig. 3. ICTD and ICLD are defined between the reference channel 1 and each of the other $C-1$ channels.

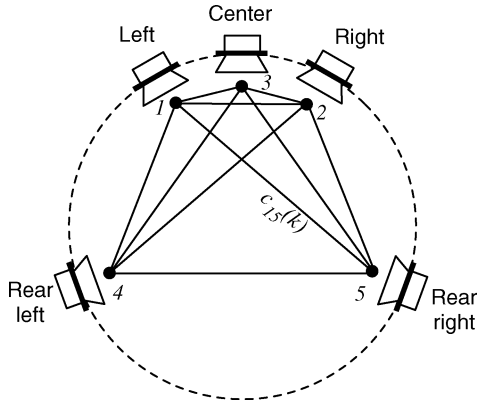

Fig. 4. In the most general case, ICC is considered between each possible channel pair.

- ICTD [samples]

$$\tau_{12}(k) = \arg\max_d \{\Phi_{12}(d,k)\} \tag{2}$$

with a short-time estimate of the normalized cross-correlation function

$$\Phi_{12}(d,k) = \frac{p_{\tilde{x}_1\tilde{x}_2}(d,k)}{\sqrt{p_{\tilde{x}_1}(k-d_1)p_{\tilde{x}_2}(k-d_2)}} \tag{3}$$

where

$$\begin{aligned} d_1 &= \max\{-d, 0\} \\ d_2 &= \max\{d, 0\} \end{aligned} \tag{4}$$

and $p_{\tilde{x}_1\tilde{x}_2}(d,k)$ is a short-time estimate of the mean of $\tilde{x}_1(k-d_1)\tilde{x}_2(k-d_2)$.

- ICC

$$c_{12}(k) = \max_d |\Phi_{12}(d,k)|. \tag{5}$$

Note that the absolute value of the normalized cross correlation is considered and $c_{12}(k)$ has a range of [0, 1].

For stereo audio signals, ICTD, ICLD, and ICC are defined between the left and right signal channel. For multichannel audio signals, it is enough to define ICTD and ICLD between a reference channel (e.g., channel number 1) and the other channels [5], as illustrated in Fig. 3 for the case of $C = 5$ channels. $\tau_{1i}(k)$ and $\Delta L_{1i}(k)$ denote the ICTD and ICLD between the reference channel 1 and channel i.

As opposed to ICTD and ICLD, ICC has more degrees of freedom. The ICC as defined can have different values between all possible input channel pairs. For C channels there are $C(C-1)/2$ possible channel pairs, e.g., for five channels, there are ten channel pairs, as illustrated in Fig. 4. However, such a scheme requires that for each subband at each time index $C(C-1)/2$ ICC are estimated and transmitted, resulting in high computational complexity and high bitrate.

Only one single ICC parameter per subband is used to describe the overall coherence between all audio channels. We obtained good results by estimating and transmitting only ICC cues between the two channels with most energy in each subband at each time index. This is illustrated in Fig. 5, when, for time instants $k-1$ and k, the channel pairs (3, 4) and (1, 2) are strongest, respectively. A heuristic rule is used for determining ICC between the other channel pairs (see Section V-C for details).

Fig. 6 shows conceptually how the BCC decoder generates a multichannel audio signal given the transmitted sum signal. The

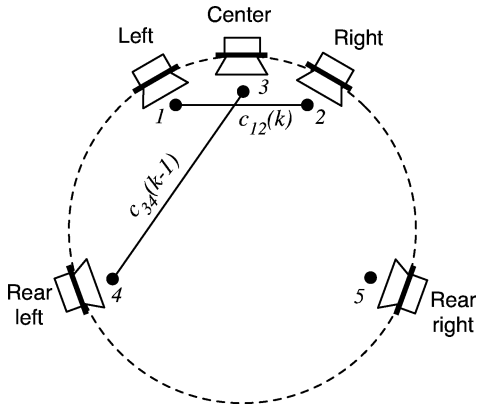

Fig. 5. At each time instant k, the ICC between the channel pair with the most power is considered. In the example shown, the channel pair is (3, 4) at time instance $k-1$ and (1, 2) at time instance k.

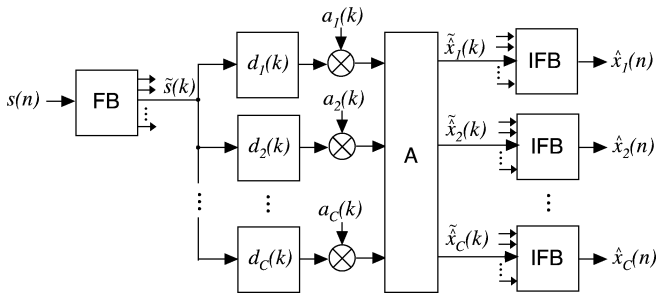

Fig. 6. ICTD are synthesized by imposing delays, ICLD by scaling, and ICC by other processing (processing block A). The shown processing is applied independently to each subband.

sum signal is decomposed into a number of subbands. The corresponding subbands for the output audio channels are generated by imposing delays (ICTD synthesis) and applying scale factors (ICLD synthesis). It is less obvious how to synthesize ICC. Several approaches for synthesizing ICC were proposed previously [5], [8]. In [5], ICC synthesis based on decorrelation by means of ICLD variation in auditory critical bands was described. In [8], an allpass filter is used to generate a second audio channel given the sum signal which is then used for stereo ICC synthesis. These techniques are computationally efficient and perform reasonably well for stereo audio signals. However, for more audio channels ICLD variation and allpass filtering as in [8] does not yield enough decorrelation capability to generate independent signal components for each audio channel. This problem is more pronounced at low frequencies (e.g., up to 1 kHz).

Both of these drawbacks are addressed by the ICC synthesis technique described in this paper. The method is scalable for any number of audio channels and effectively decorrelates even very low frequency signal components if necessary. Furthermore, the proposed scheme is able to reproduce approximately the same ICCs (maximum of the normalized cross-correlation function) as are present in the original signal, in contrast to heuristic schemes [5], [8] which are not directly based on synthesizing specific ICC values.

III. Spatial Hearing and Spatial Audio Playback

Similarly to the way humans perceive a visual image, humans are also able to perceive an *auditory spatial image*. The different objects which are part of the auditory spatial image are denoted *auditory events*. When stereo or multichannel audio signals are played back over headphones or loudspeakers they evoke an auditory spatial image in the listener. In the following, spatial hearing is discussed with emphasis on phenomena relevant for spatial audio playback.

A. Spatial Hearing With One Sound Source

The simplest listening scenario is when there is one sound source in *free field*. In this case, the ear input signals can be viewed as being filtered versions of the source signal. The filters modeling the path of sound from a source to the left and right ear entrances are commonly referred to as *head-related transfer functions* (HRTFs) [9]. For each source direction different HRTFs need to be used for modeling the ear entrance signals.

A more intuitive but only approximately valid view for the relation between the source azimuth ϕ and the ear entrance signals considers the difference in length of the paths from the source to the two ear entrances as a function of the source angle ϕ [9]. As a result of the different path lengths, there is a difference in arrival time between both ear entrances. Due to this path length difference, there is a difference in arrival times of sound at the left and right ears, denoted *interaural time difference* (ITD). Additionally, the shadowing of the head results in an intensity difference of the left and right ear entrance signals, denoted *interaural level difference* (ILD). For example, a source to the left of a listener results in a higher intensity of the signal at the left ear than at the right ear. For ITD and ILD the same definitions as for the previously defined ICTD and ICLD can be used.

Diffraction, reflection, and resonance effects caused by the head, torso, and the external ears of the listener result in that ITD and ILD not only depend on the source angle ϕ, but also on the source signal. Nevertheless, if ITD and ILD are considered as a function of frequency, it is a reasonable approximation to say that the source angle solely determines ITD and ILD as implied by data shown in [10]. When only considering frontal directions ($-90^\circ \leq \phi \leq 90^\circ$) the source angle ϕ approximately causally determines ITD and ILD. However, for each frontal direction, there is a corresponding direction in the back of the listener resulting in a similar ITD–ILD pair. Thus, the auditory system needs to rely on other cues for resolving this front/back ambiguity. Examples of such cues are head movement cues, visual cues, and spectral cues (different frequencies are emphasized or attenuated when a source is in the front or back) [9]. The following discussion does not cover these other cues, since these are not considered explicitly in BCC. For audio playback systems with loudspeakers these other cues are automatically inherent in the ear entrance signals due to the physical location of the loudspeakers.

B. Ear Entrance Signal Properties and Lateralization

Fig. 7(a) illustrates perceived auditory events for different ITD and ILD [9] for two coherent left and right headphone signals. When left and right headphone signals are coherent, have the same level (ILD $= 0$), and no delay difference (ITD $= 0$), an auditory event appears in the center between the left and right ears of a listener. More specifically, the auditory event appears in

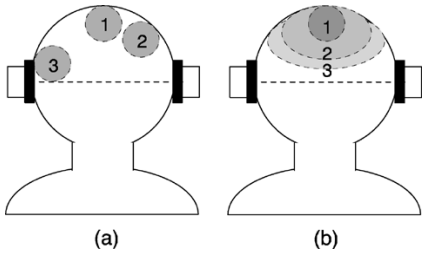

Fig. 7. (a) ILD and ITD between a pair of headphone signals determine the location of the auditory event which appears in the frontal section of the upper head. (b) The width of the auditory event increases (1–3) as the interaural coherence (IC) between the left and right headphone signals decreases.

the center of the frontal section of the upper half of the head of a listener, as illustrated by Region 1 in Fig. 7(a). By increasing the level on one side, e.g., right, the auditory event moves to that side as illustrated by Region 2 in Fig. 7(a). In the extreme case, when only the signal on the left is active, the auditory event appears at the left side as illustrated by Region 3 in Fig. 7(a). ITD can be used similarly to control the position of the auditory event.

Another ear entrance signal property that is considered in this discussion is a measure for the degree of "similarity" between the left and right ear entrance signals, denoted *interaural coherence* (IC), defined similarly as the previously defined ICC.

When two identical signals (IC $= 1$) are emitted by the two transducers of the headphones, a relatively compact auditory event is perceived. For noise the width of the auditory event increases as the IC between the headphone signals decreases, as illustrated in Fig. 7(b) [11].

C. Two Sound Sources: Summing Localization

For two sources at a distance (e.g., loudspeaker pair), ITD, ILD, and IC are determined by the HRTFs of both sources and by the specific source signals. Nevertheless, it is interesting to assess the effect of cues similar to ITD, ILD, and IC, but relative to the source signals and not ear entrance signals. To distinguish between these same properties considered either between the two ear entrance signals or two source signals, respectively, the latter are denoted ICTD, ICLD, and ICC. For headphone playback, ITD, ILD, and IC are (ideally) the same as ICTD, ICLD, and ICC. In the following a few phenomena related to ICTD, ICLD, and ICC are reviewed for two sources located in the front of a listener.

Fig. 8(a) illustrates the location of the perceived auditory events for different ICLD for two coherent source signals [9]. When left and right source signals are coherent (ICC $= 1$), have the same level (ICLD $= 0$), and no delay difference (ICTD $= 0$), an auditory event appears in the center between the two sources as illustrated by Region 1 in Fig. 8(a). By increasing the level on one side, e.g., right, the auditory event moves to that side as illustrated by Region 2 in Fig. 8(a). In the extreme case, when only the signal on the left is active, the auditory event appears at the left source position as is illustrated by Region 3 in Fig. 8(b). ICTD can be used similarly to control the position of the auditory event. This principle of controlling the location of an auditory event between a source pair is also applicable when the source pair is not in the front of the listener.

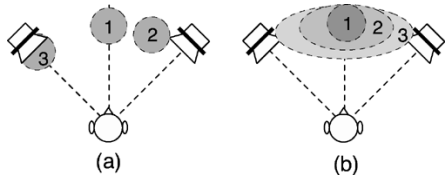

Fig. 8. (a) ICTD and ICLD between a pair of coherent source signals determine the location of the auditory event which appears between the two sources. (b) The width of the auditory event increases (1–3) as the IC between left and right source signals decreases.

However, some restrictions apply for sources to the sides of a listener [12], [13].

When coherent wideband noise signals (ICC $= 1$) are simultaneously emitted by a pair of sources, a relatively compact auditory event is perceived. When the ICC is reduced between these signals, the width of the auditory event increases [9], as illustrated in Fig. 8(b).

The insight that when signals with specific properties are emitted by two sources the direction of the auditory event can be controlled is of high relevance for applications. It is this property, which makes stereo audio playback possible. With two appropriately placed loudspeakers, the illusion of auditory events at any direction between the two loudspeakers can be generated.

Another relevance of the described phenomena is that for loudspeaker playback and headphone playback similar cues can be used for controlling the location of an auditory event. This is the basis, which makes it possible to generate signal pairs which evoke related illusions in terms of relative auditory event location for both loudspeaker and headphone playback. If this were not the case, there would be a need for different signals depending on whether a listener uses loudspeakers or headphones.

D. Other Spatial Attributes

So far, the discussion mostly focused on the attribute of perceived direction or lateralization of auditory events. One exception was the discussion of the role IC and ICC play for noise signals in determining the width of the auditory event. In the following, other attributes related to auditory events and the auditory spatial image are briefly discussed. These attributes mostly depend on the properties of reflections relative to the direct sound.

Spatial impression is defined as the impression a listener spontaneously gets about type, size, and other properties of an actual or simulated space [9]. Spatial impression is largely determined by the relation between direct sounds and reflections, and number, strength, and directions of reflections. In the following, attributes related to spatial impression are briefly reviewed. More complete reviews are given in [9] and [14].

- *Coloration:* The first early reflections up to about 20 ms later than the direct sound can cause timbral coloration due to a "comb filter" effect which attenuates and amplifies frequency components in a frequency-periodic pattern.
- *Distance of Auditory Event:* In free field, the following two ear entrance signal attributes change as a function of source distance: Power of signal reaching the ears and high frequency content (air absorption). For a source for

which a listener knows its likely level of emitted sound, such as speech, the overall sound level at the ear entrances provides an absolute distance cue [15], [16]. However, in situations when a listener does not expect a source to have a certain emitting level, overall sound level at the ear entrances can not be used for judging absolute distance [17].

On the other hand, in a reverberant environment, there is more information available to the auditory system. The reverberation time and the timing of the first reflections contain information about the size of a space and the distance to the surfaces, thus giving an indication about the expected range of source distances. For relatively distant sources the ratio of the power of direct to reflected sound is a reliable distance cue, see, e.g., [15], [16], [18].

- *Width of Auditory Events and Envelopment:* As implied by the results presented in Sections III-B and III-C, IC and ICC are related to the width of auditory events. IC can be related to the width of auditory events and *listener envelopment* [19], [20] by computing it for the early and late part of *binaural room impulse responses* (BRIRs) (e.g., up to 80 ms and later part), respectively. These two measures are often denoted early and late *interaural cross-correlation coefficient* (IACC) [21], [22]. A thorough review of IACC and related measures is given in [14].

  Since IC and ICC are in many cases directly related, i.e., lower ICC between a loudspeaker pair results in lower IC between the ear entrance signals [23], also ICC can be related to the width of auditory events and listener envelopment.

IV. Motivation for BCC to Consider ICTD, ICLD, and ICC

Given the sum signal, BCC synthesizes a stereo or multichannel audio signal such that ICTD, ICLD, and ICC approximate the corresponding cues of the original audio signal. In the following, the role of ICTD, ICLD, and ICC in relation to auditory spatial image attributes is discussed.

The discussion in Section III implies that for one auditory event ICTD and ICLD are related to perceived direction. When considering BRIRs of one source, there is a relationship between the width of the auditory event and listener envelopment and IC estimated for the early and late parts of the BRIRs. However, the relationship between IC (or ICC) and these properties for general signals (and not just the BRIRs) is not straightforward.

Stereo and multichannel audio signals usually contain a complex mix of concurrently active source signals superimposed by reflected signal components resulting from recording in enclosed spaces or added by the recording engineer for artificially creating a spatial impression. Different source signals and their reflections occupy different regions in the time-frequency plane. This is reflected by ICTD, ICLD, and ICC which vary as a function of time and frequency. In this case, the relation between instantaneous ICTD, ICLD, and ICC and auditory event directions and spatial impression is not obvious. The strategy of BCC is to simply synthesize these cues such that they approximate the corresponding cues of the original audio signal.

BCC usually uses filterbanks with subbands of bandwidths equal to two times the *equivalent rectangular bandwidth* (ERB) [24]. Informal listening revealed that the audio quality of BCC did not improve notably when choosing a higher frequency resolution. A lower frequency resolution is favorable since it results in less ICTD, ICLD, and ICC values that need to be transmitted to the decoder and, thus, in a lower bitrate.

Regarding time resolution, ICTD, ICLD, and ICC are considered at regular time intervals. Best performance is obtained when ICTD, ICLD, and ICC are considered about every 4–16 ms. Other schemes have also used time varying rates for cue synthesis [8], [25]. Note that, unless the cues are considered at very short time intervals, the *precedence effect* [9], [26] is not directly considered. Assuming a classical lead-lag pair of sound stimuli, when the lead and lag fall into a time interval where only one set of cues is synthesized, localization dominance of the lead is not considered. Despite of this, BCC achieves good audio quality on average and up to nearly transparent quality for certain audio signals.

The often-achieved perceptually small difference between reference signal and synthesized signal implies that cues related to a wide range of auditory spatial image attributes are implicitly considered by synthesizing ICTD, ICLD, and ICC at regular time intervals. In the following, some arguments are given on how ICTD, ICLD, and ICC may relate to a range of auditory spatial image attributes.

Early reflections up to about 20 ms result in coloration of sources' signals. This coloration effect is different for each audio channel determined by the timing of the early reflections contained in the channel. BCC does not attempt to retrieve the corresponding early reflected sound for each audio channel (which is a source separation problem). However, frequency dependent ICLD synthesis imposes on each output channel the spectral envelope of the original audio signal and, thus, is able to mimic coloration effects caused by early reflections.

Most perceptual phenomena related to spatial impression seem to be related directly to the nature of reflections that occur following the direct sound. This includes the nature of early reflections up to 80 ms and late reflections beyond 80 ms. Thus, it is crucial that the effect of these reflections is mimicked by the synthesized signal.

ICTD and ICLD synthesis ideally result in that each channel of the synthesized output signal has the same temporal and spectral envelope as the original signal. This includes the decay of reverberation (the sum of all reflections is preserved in the transmitted sum signal and ICLD synthesis imposes the desired decay for each audio channel individually). ICC synthesis decorrelates signal components that were originally decorrelated by lateral reflections. Also, there is no need of considering reverberation time explicitly. Blindly synthesizing ICC at each time instant to approximate ICC of the original signal has the desired effect of mimicking different reverberation times, since ICLD synthesis imposes the desired rate of decay.

The most important cues for auditory event distance are overall sound level and direct sound to total reflected sound ratio [27]. Since BCC generates level information and reverberation such that it approaches that of the original signal, also auditory event distance cues are represented by considering ICTD, ICLD, and ICC cues.

V. Synthesis of ICTD, ICLD, and ICC

A. Generating Decorrelated Audio Channels

A natural listening scenario where low interaural coherence (IC) occurs is in a concert hall. Late reverberation sound arrives at the ears from random angles with random strength, such that the IC is low. Particularly, lateral reflections, known to be important for good concert hall acoustics, result in low IC [28]. The property of late reverberation to result in low IC gives the motivation for generating a number of decorrelated audio channels by applying late reverberation models to the given transmitted BCC sum signal $s(n)$.

C late reverberation audio channels $s_c(n)$ $(1 \leq c \leq C)$ with low ICC between pairs of channels are obtained by

$$s_c(n) = h_c(n) \star s(n) \tag{6}$$

where $\star$ denotes convolution and $h_i(n)$ are the filters modeling late reverberation. Late reverberation is modeled by impulse responses given by

$$h_i(n) = \begin{cases} r_i(n)\left(1 - \frac{1}{f_s T_h}\right)^n, & 0 \leq n < M \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where $r_i(n)$ $(1 \leq i \leq C)$ are independent stationary white Gaussian noise signals, T_h is the time constant in seconds of the exponential decay of the impulse response, f_s is the sampling frequency, and M is the length of the impulse response in samples. An exponential decay is chosen because the strength of late reverberation is decaying exponentially in time.

The reverberation time of many concert halls is in the range of 1.5–3.5 s [29]. In order that the late reverberation signals $s_c(n)$ are enough independent for generating as low IC or ICC as occur in concert halls, T_h is chosen such that reverberation times of $h_c(n)$ are in the same range. Given an impulse response, the reverberation time is computed as described in [30]. We choose $T_h = 0.4$ s, corresponding to a reverberation time of 2.8 s. M is chosen such that the impulse responses are 0.3 s long (note that the reverberation time considerations were only used to determine the rate of decay and not the length of the filters; the length of the filters was determined heuristically to be as short as possible without compromising the quality of the BCC synthesis scheme).

In the following, we describe how the late reverberation audio channels $s_c(n)$ are used to synthesize ICC. Each subband of each output signal channel is computed as a weighted sum of the corresponding subbands of $s(n)$ and $s_c(n)$ $(1 \leq c \leq C)$. The factors of the weighted sum are determined such that the ICC cues between the output subbands approximate those of the original audio signal.

As mentioned previously, ICC are synthesized every 4–16 ms. The length and decay of the filters h_c determines the upper bound of decorrelation capability of the system. In Section IV it was discussed how by spatial cue synthesis different degrees of reverberation are synthesized. Since the spatial cues are updated every 4–16 ms, the decorrelation effect of the long tail of h_c is effectively suppressed if necessary. Thus, the long impulse responses are suitable not only for synthesizing ICC mimicking very reverberant environments, but also for ICC for mimicking less reverberant environments.

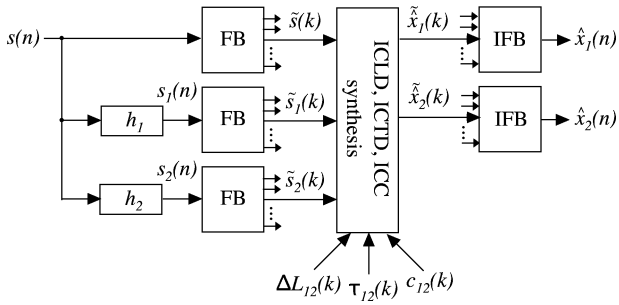

Fig. 9. BCC synthesis scheme for generating stereo signals.

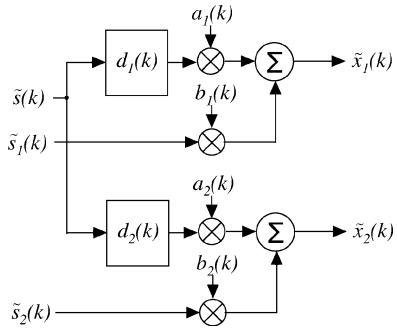

Fig. 10. Processing of one subband for ICTD, ICLD, and ICC synthesis for stereo.

B. Stereo BCC Synthesis

The proposed scheme is shown in Fig. 9. The late reverberation channels $s_1(n)$ and $s_2(n)$ are generated by filtering the sum signal $s(n)$ as described previously. The signals $s(n)$, $s_1(n)$, and $s_2(n)$ are decomposed into a number of spectral components by nonuniform filterbanks with subbands reflecting the frequency resolution of the auditory system. One subband signal of the decomposed three input signals is denoted $\tilde{s}(k)$, $\tilde{s}_1(k)$, or $\tilde{s}_2(k)$, respectively. As illustrated in Fig. 10, the subbands of the two output channels are computed as a weighted sum of the given subband signals

$$\begin{aligned} \tilde{\tilde{x}}_1(k) &= a_1 \tilde{s}(k - d_1) + b_1 \tilde{s}_1(k) \\ \tilde{\tilde{x}}_2(k) &= a_2 \tilde{s}(k - d_2) + b_2 \tilde{s}_2(k) \end{aligned} \tag{8}$$

where the scale factors (a_1, a_2, b_1, b_2) and delays (d_1, d_2) are determined as a function of the desired ICTD $\tau_{12}(k)$, ICLD $\Delta L_{12}(k)$, and ICC $c_{12}(k)$ (the time index of the gain factors and delays is neglected for a simpler notation). The subband signals $\tilde{\tilde{x}}_1(k)$ and $\tilde{\tilde{x}}_2(k)$ are computed for all subbands and the output signals $x_1(n)$ and $x_2(n)$ are generated by applying the inverse of the filterbank used.

The ICTD $\tau_{12}(k)$ is synthesized by imposing two different delays, d_1 and d_2, on $\tilde{s}(k)$. These delays are computed by (4) with $d = \tau_{12}(n)$. If the subband sampling rate does not provide high enough time resolution for ICTD synthesis, delays can be imposed more precisely by using suitable all-pass filters. In

order that the output subband signals have an ICLD (1) equal to $\Delta L_{12}(k)$, the gain factors (a_1, a_2, b_1, b_2) must satisfy

$$\frac{a_2^2 p_{\tilde{s}}(k) + b_2^2 p_{\tilde{s}_2}(k)}{a_1^2 p_{\tilde{s}}(k) + b_1^2 p_{\tilde{s}_1}(k)} = 10^{\Delta L_{12}(k)/10} \tag{9}$$

where $p_{\tilde{s}}(k)$, $p_{\tilde{s}_1}(k)$, and $p_{\tilde{s}_2}(k)$ are short-time power estimates of the subband signals $\tilde{s}(k)$, $\tilde{s}_1(k)$, and $\tilde{s}_2(k)$, respectively.

For the output subband signals to have a certain ICC (5), $c_{12}(k)$, the gain factors must satisfy

$$\frac{a_1 a_2 p_{\tilde{s}}(k)}{\sqrt{(a_1^2 p_{\tilde{s}}(k) + b_1^2 p_{\tilde{s}_1}(k))(a_2^2 p_{\tilde{s}}(k) + b_2^2 p_{\tilde{s}_2}(k))}} = c_{12}(k) \tag{10}$$

as is easily shown by applying (3) and (5) to the signals given in (8) and assuming that $\tilde{s}(k)$, $\tilde{s}_1(k)$, and $\tilde{s}_2(k)$ are independent.

BCC synthesis usually normalizes its output signals, such that the sum of the power of all output channels is equal to the power of the input sum signal [5]. This yields another equation for the gain factors

$$(a_1^2 + a_2^2) p_{\tilde{s}}(k) + b_1^2 p_{\tilde{s}_1}(k) + b_2^2 p_{\tilde{s}_2}(k) = p_{\tilde{s}}(k). \tag{11}$$

Since there are four gain factors and three equations, there is one degree of freedom in the choice of the gain factors. Thus, an additional condition can be formulated

$$b_1^2 p_{\tilde{s}_1}(k) = b_2^2 p_{\tilde{s}_2}(k). \tag{12}$$

Condition (12) forces the amount of late reverberation to be the same in the left and right channel. There are several motivations for doing this.

- Late reverberation as appearing in concert halls has a level which is nearly independent of position (for relatively small displacements). Thus, the level difference of the late reverberation between left and right is always about 0 dB.
- The sound of the stronger channel is modified less, reducing negative effects of the long convolutions (6), such as time spreading of transients.

The gain factors are the nonnegative solutions of the equation system given by (9)–(12)

$$\begin{aligned} a_1 &= \sqrt{\frac{1 - A + B}{C}} \\ a_2 &= \sqrt{\frac{A - 1 + B}{C}} \\ b_1 &= \sqrt{\frac{(A + 1 - B) p_{\tilde{s}}(k)}{C p_{\tilde{s}_1}(k)}} \\ b_2 &= \sqrt{\frac{(A + 1 - B) p_{\tilde{s}}(k)}{C p_{\tilde{s}_2}(k)}} \end{aligned} \tag{13}$$

with

$$\begin{aligned} A &= 10^{\Delta L_{12}(k)/10} \\ B &= \sqrt{(1 - A)^2 + 4 A c_{12}^2(k)} \\ C &= 2(1 + A). \end{aligned} \tag{14}$$

C. Multichannel BCC Synthesis

Similar to the stereo case, the output subband signals are computed as weighted sums of the subband signals of the sum signal and diffuse audio channels

$$\begin{aligned} \tilde{x}_1(k) &= a_1 \tilde{s}(k - d_1) + b_1 \tilde{s}_1(k) \\ \tilde{x}_2(k) &= a_2 \tilde{s}(k - d_2) + b_2 \tilde{s}_2(k) \\ &\vdots \quad \vdots \\ \tilde{x}_C(k) &= a_C \tilde{s}(k - d_C) + b_C \tilde{s}_C(k) \end{aligned} \tag{15}$$

as is illustrated in Fig. 12: The delays are determined by the ICTDs

$$d_i = \begin{cases} -\min_{1 \le l < C} \tau_{1l}(k), & i = 1 \\ \tau_{1i}(k) + d_1, & 2 \le i \le C. \end{cases} \tag{16}$$

$2C$ equations are needed to determine the $2C$ scale factors in (15). In the following the conditions leading to these equations are briefly described.

- *ICLD:* $C - 1$ equations similar to (9) are formulated between the channel pairs such that the output subband signals have the desired ICLD cues.
- *ICC for the two strongest channels:* Two equations similar to (10) and (12) between the two strongest audio channels, i_1 and i_2, are formulated such that the ICC between these channels is the same as estimated in the encoder and the amount of diffuse sound in both channels is the same, respectively.
- *Normalization:* One equation is obtained by extending (11) to C channels

$$\sum_{i=1}^{C} a_i^2 p_{\tilde{s}}(k) + \sum_{i=1}^{C} b_i^2 p_{\tilde{s}_i}(k) = p_{\tilde{s}}(k). \tag{17}$$

- *ICC for $C - 2$ weakest channels:* The ratio between the power of diffuse sound to nondiffuse sound for the weakest $C - 2$ channels ($i \neq i_1 \wedge i \neq i_2$) is chosen to be the same as for the second strongest channel i_2

$$\frac{b_i^2 p_{\tilde{s}_i}(k)}{a_i^2 p_{\tilde{s}}(k)} = \frac{b_{i_2}^2 p_{\tilde{s}_{i_2}}(k)}{a_{i_2}^2 p_{\tilde{s}}(k)} \tag{18}$$

resulting in another $C - 2$ equations, for a total of $2C$ equations. The gain factors are the nonnegative solutions of the described $2C$ equations.

VI. Reducing Computational Complexity

As mentioned before, for reproducing naturally sounding diffuse sound [31], the impulse responses $h_i(n)$ (7) need to be as long as several hundred milliseconds, resulting in high computational complexity. Furthermore, BCC synthesis requires for each $h_i(n)$ $(1 \le i \le C)$ additional filterbank processing (Figs. 9 and 11) .

The computational complexity could be reduced by using artificial reverberation algorithms [32], [33] for generating late reverberation and using that for $s_i(n)$. Another possibility is to still carry out the convolutions (6) but applying an algorithm

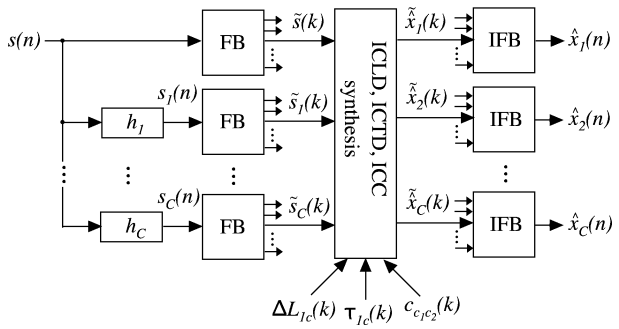

Fig. 11. BCC synthesis scheme for generating multichannel audio signals.

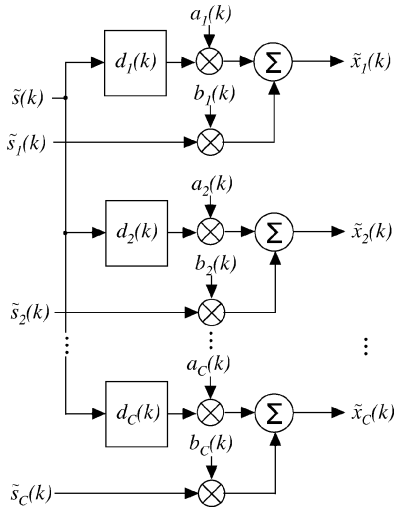

Fig. 12. Processing of one subband for ICTD, ICLD, and ICC synthesis for multichannel.

based on the *fast Fourier transform* (FFT) for reduced computational complexity [34]. We chose a related method. The difference to [34] is, that we are operating in the (discrete) *short time Fourier transform* (STFT) domain, i.e., we are using overlapping windows. The motivation is to use the same STFT for carrying out the convolutions and the BCC processing [5]. This results in lower computational complexity of the convolution computation and no need in using an additional filterbank for each $h_i(n)$. The technique is derived for a generic signal and impulse response, denoted $x(n)$ and $h(n)$, respectively.

The STFT applies *discrete Fourier transforms* (DFTs) to windowed portions of a signal $x(n)$. The windowing is applied at regular intervals, denoted window hop size N. The resulting windowed signal with window position index k is

$$x_k(n) = \begin{cases} w(n-kN)x(n), & kN \le n < kN+W \\ 0, & \text{otherwise} \end{cases} \quad (19)$$

where W is the window length. We are using a Hann window of length $W = 512$ samples and a window hop size of $N = W/2$ samples. Other windows can be used which fulfill the (in the following assumed) condition:

$$x(n) = \sum_{k=-\infty}^{\infty} x_k(n). \quad (20)$$

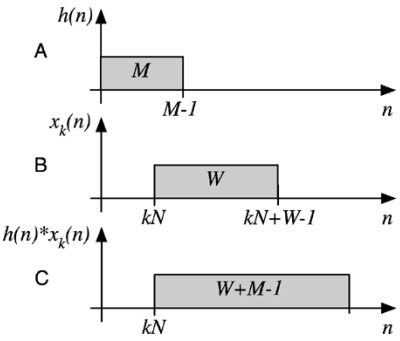

Fig. 13. Illustration of the nonzero span of $h(n)$, $x_k(n)$, and $h(n) \star x_k(n)$.

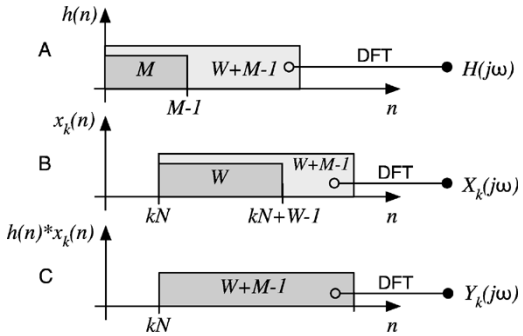

Fig. 14. Illustration of the DFTs of size $W + M - 1$ that are applied to $h(n)$, $x_k(n)$, and $h(n) \star x_k(n)$.

First, the simple case of implementing a convolution of the windowed signal $x_k(n)$ in the frequency domain is considered. Panel A of Fig. 13 illustrates the nonzero span of an impulse response $h(n)$ of length M. Similarly, the nonzero span of $x_k(k)$ is illustrated in Panel B. It is easy to verify that $h(n) \star x_k(k)$ has a nonzero span of $W + M - 1$ samples as illustrated in Panel C.

Fig. 14 illustrates at which time indexes DFTs of length $W + M - 1$ are applied to the signals $h(n)$, $x_k(n)$, and $h(n) \star x_k(k)$, respectively. Panel A of Fig. 14 illustrates that $H(j\omega)$ denotes the spectrum obtained by applying the DFT starting at time index $n = 0$ to $h(n)$. Panels B and C of Fig. 14 illustrate the computation of $X_k(j\omega)$ and $Y_k(j\omega)$ from $x_k(n)$ and $h(n) \star x_k(n)$, respectively, by applying the DFTs starting at time index $n = kN$. Note that the DFT spectra are discrete. Despite of this the notation here uses a continuous frequency variable ω. It can easily be shown that $Y_k(j\omega) = H(j\omega)X_k(j\omega)$. That is, because the zeros at the end of the signals $h(n)$ and $x_k(n)$ result in that the circular convolution imposed on the signals by the spectrum product is equal to linear convolution.

From the linearity property of convolution and (20), it follows that:

$$h(n) \star x(n) = \sum_{k=-\infty}^{\infty} h(n) \star x_k(n). \quad (21)$$

Thus, it is possible to implement a convolution in the domain of the STFT, by computing at each time k the product $H(j\omega)X_k(j\omega)$ and applying the inverse STFT (inverse DFT plus overlap/add). [A DFT of length $W + M - 1$ (or longer) needs to be used with zero padding as implied by Fig. 14]. The described technique is similar to overlap/add convolution [35]

with the generalization that overlapping windows can be used [any window fulfilling condition (20)].

The described method is not practical for long impulse responses (e.g., $M \gg W$) since then a DFT of a much larger size than W needs to be used. In the following, we are extending the described method such that only a DFT of size $W+N-1$ needs to be used.

A long impulse response $h(n)$ of length $M = LN$ is partitioned into L shorter impulse responses

$$h_l(n) = \begin{cases} h(n+lN), & 0 \le n < N \\ 0, & \text{otherwise.} \end{cases} \tag{22}$$

[If $\mathrm{mod}(M,N) \neq 0$ then $N - \mathrm{mod}(M,N)$ zeroes are added to the tail of $h(n)$]. The convolution with $h(n)$ can then be written as a sum of shorter convolutions

$$h(n) \star x(n) = \sum_{l=0}^{L-1} h_l(n) \star x(n-lN). \tag{23}$$

Applying (21) and (23) at the same time yields

$$h(n) \star x(n) = \sum_{k=-\infty}^{\infty} \sum_{l=0}^{L-1} h_l(n) \star x_k(n-lN). \tag{24}$$

The nonzero time span of one convolution in (24), $h_l(n) \star x_k(n-lN)$, as a function of k and l is $(k+l)N \le n < (k+l+1)N+W$. Thus, for obtaining its spectrum, $\tilde{Y}_{kl}(j\omega)$, the DFT is applied to this interval (corresponds to DFT position index $k+l$). It can easily be shown that $\tilde{Y}_{kl}(j\omega) = H_l(j\omega)X_k(j\omega)$, where $X_k(j\omega)$ is defined as previously with $M = N$ and $H_l(j\omega)$ is defined similarly as previously $H(j\omega)$ but for the impulse response $h_l(n)$.

The sum of all spectra $\tilde{Y}_{kl}(j\omega)$ with the same DFT position index $i = k + l$ is

$$Y_i(j\omega) = \sum_{i=0}^{k} \sum_{l=0}^{k} q(i,l)\tilde{Y}_{i+l}(j\omega) \tag{25}$$

where $q(i,l) = 1$ if $i + l = k$ and $q(i,l) = 0$ otherwise. This can be reformulated as

$$Y_k(j\omega) = \sum_{l=0}^{L-1} H_l(j\omega)X_{k-l}(j\omega). \tag{26}$$

Thus, the convolution $h(n) \star x_k(n)$ is implemented in the STFT domain by applying (26) at each spectrum index i to obtain $Y_i(j\omega)$. The inverse STFT (inverse DFT plus overlap/add) applied to $Y_i(j\omega)$ is equal to the convolution $h(n) \star x(n)$, as desired.

Note that independently of the length of $h(n)$, the required amount of zero padding is upper bounded by $N-1$ (one sample less than the STFT window hop size). DFTs larger than $W + N - 1$ can be used if desired (e.g., using an FFT with a length equal to a power of two).

As mentioned before, low complexity BCC synthesis operates also in the STFT domain. In this case, ICTD, ICLD, and ICC synthesis is applied to groups of STFT bins representing spectral components with bandwidths equal or proportional to the bandwidth of a critical band as described in [5] (where groups of bins are denoted "partitions"). In such a system, for reduced complexity, the inverse STFT is not applied to (26), but the spectra of (26) are directly used as diffuse sound in the frequency domain.

VII. Subjective Evaluation

A. Subjects and Playback Setup

The test was conducted in two different listening rooms with different equipment and subjects at different locations (EPFL Lausanne, Fraunhofer IIS).

1) Four adults with an age range of 22–29 participated as subjects in the listening tests at EPFL. Two subjects are experienced listeners and two are nonexperienced. During the test, the subjects were sitting on a chair that was placed in the sweetspot of a standard 5.1 listening setup [36] in a sound insulated room. The loudspeakers were placed on a circle with a radius of 2 m. Informal listening revealed that a relatively small scale loudspeaker setup is more critical than a larger scale loudspeaker setup. For audio playback, an *Apple PowerBook G4* laptop computer was used with an external multichannel D/A converter (*Emagic EMI A26*) directly connected to active loudspeakers (*Genelec 1031A*).
2) Five experienced adult listeners participated in the listening test at Fraunhofer IIS. During the test, the subjects were sitting on a chair that was placed in the sweetspot of a standard 5.1 listening setup in a sound insulated room. A personal computer with an *RME Hammerfall* digital sound output interface connected to *Lake People DAC F20* D/A converters was used. The D/A converters were directly connected to active loudspeakers (*Geithain RL 901*).

B. Stimuli

Different kinds of reference five–channel audio material was selected: Classical recordings mimicking a concert hall experience and movie soundtrack style items with auditory events occurring in all directions. We chose audio material that we consider critical for multichannel BCC coding (e.g., applause). The reference items (R) were compared to two kinds of BCC synthesized items: BCC with ICC synthesis (A) and BCC without ICC synthesis (B). We included items B in the evaluation for assessing the improvement achieved when considering ICC over the case of not considering ICC. The sum signal was not coded to avoid affecting the test results due to coding artifacts. ICLD and ICC were transmitted to the decoder in full precision, i.e., no quantization and coding was applied. For five–channel audio material, the proposed scheme needs to transmit for each subband and time index 4 ICTD and 4 ICLD values and, if used, one ICC value. With the quantization and coding scheme proposed in [5] this would result in a bitrate for the BCC side information of $8 \cdot 2$ kb/s for ICTD and ICLD, and about 2 kb/s for ICC.

We did not compare the proposed scheme to previous schemes for ICC synthesis, since the quality of the scheme

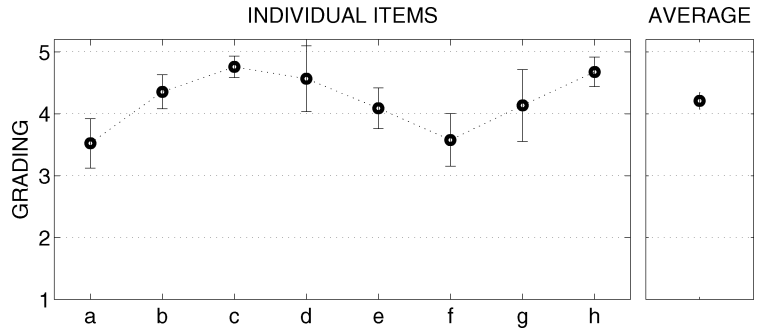

Fig. 15. Test 1: Hidden reference test results. The test results averaged over the subjects and 95% confidence intervals are shown for each item (left panel) and averaged for all items (right panel). (Grading scale judges difference between BCC with ICC synthesis and reference: 5: "not perceptible," 4: "perceptible but not annoying," 3: "slightly annoying," 2: "annoying," 1: "very annoying").

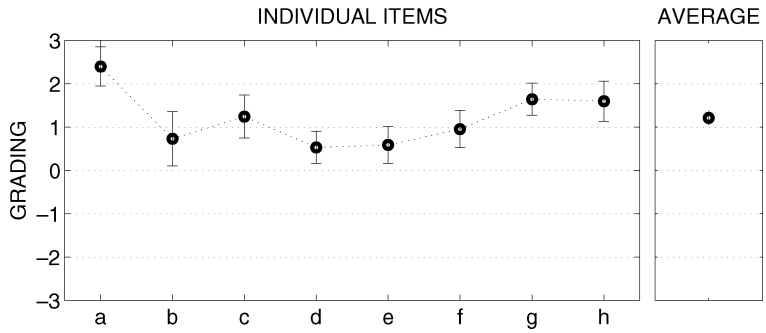

Fig. 16. Test 2: Seven-grade relative comparison scale test results. The test results averaged over the subjects and 95% confidence intervals are shown for each item (left panel) and averaged for all items (right panel). (Positive values correspond to better performance of BCC with ICC compared to BCC without ICC synthesis: 1 "slightly better," 2: "better," 3: "much better").

proposed in [5] for multichannel signals is not satisfactory and the other scheme [8] is only applicable to stereo.

C. Test Methods

1) *Test 1:* A test was conducted to assess the quality of items A (BCC with ICC synthesis) relative to the reference items R. The test method used was the hidden reference method, used according to [37]. The reference item is played, followed by the reference item and the degraded item in random order. A five–grade impairment scale was used for comparing the degraded item to the reference. After the three items were initially played, the listener could selectively listen to the items again while switching between the items at any time. This method is suitable for subjective assessment of small impairments. We decided to use this method, after informal listening revealed that for the considered items the degree of impairment is fairly small.
2) *Test 2:* Informal listening revealed that items B (BCC without ICC synthesis) were impaired significantly more than items A (BCC with ICC synthesis). Therefore, we decided not to compare these items with an absolute scale to the reference items, but to use a relative seven-grade comparison scale to compare items A and B, according to [38]. In this case, the reference item is played, followed by items A and B in random order. Similarly, as in test 1, after the three items were initially played, the listener could selectively listen to the items again while switching between the items at any time.

D. Results

The results for both tests obtained from the two different locations (EPFL Lausanne, Fraunhofer IIS) were fairly similar, and, thus, we averaged the results over all subjects at both locations.

1) *Test 1:* Fig. 15 shows the results for the individual items averaged for all subjects and the overall average. The proposed scheme for ICTD, ICLD, and ICC synthesis has an overall grading between "perceptible but not annoying" and "imperceptible." Considering that the items are synthesized from only the sum signal and the low bitrate (mono transmission plus 18 kb/s side information), this is an excellent result.

   The items with the best quality in Fig. 15 (*b*, *c*, *d*, *h*) are a classical recording, movie soundtracks, and a scene with auditory events all around the subject. The most critical item is the applause signal (*a*). Item *e* also contains critical applause and a talker at the side. Item *f* is a classical recording with very tonal components, where the ICC synthesis introduces some distortions. Item *g* is a movie soundtrack signal. The degradations of the BCC items often result from a spatial impression which is slightly different than the spatial impression of the reference items. The applause signal not only suffers from a modified spatial impression, but also from time spreading of its transients.

2) *Test 2:* The results are shown in Fig. 16 for the individual items averaged for all subjects and the overall average. The items with ICC synthesis are significantly better than the items without ICC synthesis. The proposed ICC synthesis gives an improvement for all items compared to no ICC synthesis. Often, the items without ICC synthesis sound unnatural and colored, and spatial impression is largely lost. Interestingly, the worst item in Test 1 (item *a*, applause) is most improved by ICC synthesis.

VIII. Conclusion

We proposed a scheme for stereo and multichannel synthesis of ICC cues for parametric stereo and multichannel coding. The scheme synthesizes ICC cues such that they approximate those of the original audio signal. For that purpose, diffuse audio channels are generated and mixed with the transmitted sum signal. The diffuse audio channels are generated using long filters with exponentially decaying Gaussian impulse responses. Such impulse responses generate diffuse sound similar to late reverberation. An alternative implementation for reduced computational complexity is proposed. ICTD, ICLD, and ICC synthesis are all carried out in the domain of a single STFT, including the filtering for diffuse sound generation.

The results of a subjective assessment indicate that the proposed scheme provides good audio quality. Another test indicates that the proposed ICC synthesis results in significantly improved audio quality compared to our previous scheme without ICC synthesis for each item in the test.

Acknowledgment

The author would like to thank A. Härmä, J. Herre, H. Järveläinen, M. Karjalainen, and V. Pulkki for the inspiring discussions on diffuse sound, reverberation algorithms, and multichannel amplitude panning. He would also like to thank C. Spenger for conducting the subjective test at Fraunhofer IIS and F. Baumgarte, P. Kroon, and the anonymous reviewers for their valuable suggestions for improving this manuscript.

References

[1] R. G. v. d. Waal and R. N. J. Veldhuis, "Subband coding of stereophonic digital audio signals," in *Proc. IEEE ICASSP*, 1991, pp. 3601–3604.

[2] J. Herre, K. Brandenburg, and D. Lederer, "Intensity stereo coding," presented at the 96th Convention Audio Engineering Soc.", Feb. 1994.

[3] F. Baumgarte and C. Faller, "Why binaural cue coding is better than intensity stereo coding," presented at the 112th Convention Audio Engineering Soc., May 2002.

[4] H. S. Malvar, *Signal Processing with Lapped Transforms*. Norwood, MA: Artech House, 1992.

[5] C. Faller and F. Baumgarte, "Binaural cue coding—Part II: Schemes and applications," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 520–531, Nov. 2003.

[6] ——, "Binaural cue coding: A novel and efficient representation of spatial audio," *Proc. ICASSP*, vol. 2, pp. 1841–1844, May 2002.

[7] F. Baumgarte and C. Faller, "Binaural cue coding—Part I: Psychoacoustic fundamentals and design principles," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 509–519, Nov. 2003.

[8] E. Schuijers, W. Oomen, B. den Brinker, and J. Breebaart, "Advances in parametric coding for high-quality audio," presented at the 112th Convention Audio Engineering Soc., Mar. 2003.

[9] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA: MIT Press, 1997.

[10] W. Gaik, "Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling," *J. Acoust. Soc. Amer.*, vol. 94, no. 1, pp. 98–110, Jul. 1993.

[11] R. I. Chernyak and N. A. Dubrovsky, "Pattern of the noise images and the binaural summation of loudness for the different interaural correlation of noise," in *Proc. 6th Int. Congr. Acoustics*, vol. 1, Tokyo, Japan, 1968, pp. A–3–A–12.

[12] G. Theile and G. Plenge, "Localization of lateral phantom sources," *J. Audio Eng. Soc.*, vol. 25, no. 4, pp. 196–200, 1977.

[13] V. Pulkki, "Localization of amplitude-panned sources II: Two- and three-dimensional panning," *J. Audio Eng. Soc.*, vol. 49, no. 9, pp. 753–757, 2001.

[14] R. Mason, "Elicitation and measurement of auditory spatial attributes in reproduced sound," Ph.D. dissertation, Dept. Music Sound Recording, Univ. Surrey, Surrey, U.K., 2002.

[15] D. H. Mershon and L. E. King, "Intensity and reverberation as factors in the auditory perception of egocentric distance," *Percep. Psychophys.*, vol. 18, no. 6, pp. 409–415, 1975.

[16] D. H. Mershon and J. N. Bowers, "Absolute and relative cues for the auditory perception of egocentric distance," *Perception*, vol. 8, pp. 311–322, 1979.

[17] P. D. Coleman, "Failure to localize the source distance of an unfamiliar sound," *J. Acoust. Soc. Amer.*, vol. 34, pp. 345–346, 1962.

[18] A. Bronkhorst and T. Houtgast, "Auditory distance perception in rooms," *Nature*, vol. 397, pp. 517–520, Feb. 1999.

[19] M. Morimoto and Z. Maekawa, "Auditory spaciousness and envelopment," in *Proc. 13th Int. Congr. Acoustics*, vol. 2, Belgrade, Yugoslavia, 1989, pp. 215–218.

[20] J. S. Bradley and B. A. Soulodre, "Listener envelopment: An essential part of good concert hall acoustics," *J. Acoust. Soc. Amer.*, vol. 99, p. 22, Jan. 1996.

[21] J. S. Bradley, "Comparison of concert hall measurements of spatial impression," *J. Acoust. Soc. Amer.*, vol. 96, no. 6, pp. 3525–3535, 1994.

[22] T. Okano, L. L. Beranek, and T. Hidaka, "Relations among interaural cross-correlation coefficient ($IAAC_E$), lateral fraction (LF_E), and apparent source width (ASW) in concernt halls," *J. Acoust. Soc. Amer.*, vol. 104, no. 1, pp. 255–265, Jul. 1998.

[23] K. Kurozumi and K. Ohgushi, "The relationship between the cross-correlation coefficient of two-channel acoustic signals and sound image quality), and apparent source width (ASW) in concernt halls," *J. Acoust. Soc. Amer.*, vol. 74, no. 6, pp. 1726–1733, Dec. 1983.

[24] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.*, vol. 47, pp. 103–138, 1990.

[25] E. Schuijers, W. Oomen, A. C. den Brinker, and A. J. Gerrits, "Advances parametric coding for high-quality audio," presented at the MPCA, Nov. 2002.

[26] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The precedence effect," *J. Acoust. Soc. Amer.*, vol. 106, no. 4, pp. 1633–1654, Oct. 1999.

[27] B. G. Shinn-Cunningham, "Distance cues for virtual auditory space," in *Proc. 1st IEEE Pacific-Rim Conf. Multimedia*, Sydney, Australia, Dec. 2000, pp. 227–230.

[28] H. Kuttruff, *Room Acoustics*. Barking, U.K.: Elsevier, 1991.

[29] L. L. Beranek, *Music, Acoustics and Architecture*. NewYork: Wiley, 1962.

[30] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Amer.*, vol. 37, pp. 409–412, 1965.

[31] M. Karjalainen and H. Järveläinen, "More about this reverberation science: Perceptually good late reverberance," presented at the 112th Convention Audio Engineering Soc., Nov. 2001.

[32] M. R. Schroeder, "Natural sounding artificial reverberation," *J. Aud. Eng. Soc.*, vol. 10, no. 3, pp. 219–223, 1962.

[33] W. G. Gardner, "Reverberation algorithms," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds. Norwell, MA: Kluwer, 1998, ch. 2.

[34] ——, "Efficient convolution without input-output delay," *J. Audio Eng. Soc.*, vol. 43, no. 3, pp. 127–136, 1995.

[35] A. V. Oppenheim and R. W. Schaefer, *Discrete-Time Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 1989.

[36] "Multi-channel stereophonic sound system with or without accompanying picture," ITU, Rec. ITU-R BS.775, 1993.

[37] "Methods for subjective assessment of small impairments in audio systems including multichannel surround systems," ITU, Rec. ITU-R BS.1116.1, 1997.

[38] "Subjective assessment of sound quality," ITU, Rec. ITU-R BS.562.3, 1990.

Christof Faller received the M.S. (Ing.) degree in electrical engineering from ETH Zurich, Zurich, Switzerland, in 2000 and the Ph.D. degree in computer and communication sciences from EPFL Lausanne, Lausanne, Switzerland, in 2004.

During his studies, he was an independent Consultant for Swiss Federal Labs, applying neural networks to process parameter optimization of sputtering processes, and he spent one year at the Czech Technical University (CVUT), Prague, Czech Republic. In 2000, he became a Consultant for the Speech and Acoustics Research Department, Bell Laboratories, Lucent Technologies, Murray Hill, NJ, and after one and a half years of consulting, partially in Europe, he became a Member of Technical Staff, focusing on new techniques for audio coding applied to digital satellite radio broadcasting. At the Lucent spin-off Agere Systems, he developed algorithms for parametric coding of multichannel audio signals, echo control, and other communications-related audio applications. He is currently with the Audiovisual Communications Laboratory, EPFL Lausanne.